2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

SMART (splitting-merging assisted reliable) Independent Component Analysis for Brain Functional Networks *

Yuhui Du*, Xingyu He, and Vince D Calhoun

Abstract— **Independent component analysis (ICA) has been widely applied to estimate brain functional networks from functional magnetic resonance imaging (fMRI) data. ICA is a data-driven approach, however, the number of components must be prespecified. Indeed, it is difficult to estimate or determine an optimal number of components in fMRI analysis. In this paper, we propose a SMART (splitting-merging assisted reliable) ICA to overcome the problem. Our method first estimates group-level components using different settings and then yields reliable components by using a splitting and merging clustering approach. Subject-specific components are obtained using our previously proposed group information guided ICA (GIG-ICA) based on reliable group-level components to estimate individual-subject independent components. Simulations with unique components for subjects showed our method extracted components with high similarity to the ground truth spatial maps (SMs). For real fMRI data, the functional networks extracted by our method showed both similarity and specificity across subjects. To sum up, our method can effectively and accurately identify subject-specific brain functional networks without a need of parameter setting.**

Clinical Relevance— **SMART ICA automatically extracts reliable subject-specific brain functional networks that can be used for biomarker identification.**

I. Introduction

Brain functional network analysis using functional magnetic resonance imaging (fMRI) data plays an important role in the neuroscience field, as network biomarkers are expected to benefit the diagnosis of mental disorders. Although different network analysis methods have been proposed, independent component analysis (ICA) is one of the most widely used methods and has been applied to explore functional impairments for various mental disorders [1, 2].

Relative to the region of interest (ROI) based approaches, ICA is data-driven, because there is no need of defining a priori brain regions in ICA. However, one must select or estimate the number of components, which hinders its automation ability to some extent. Ideally, the component number should be equal to the real source number. However, in practice, it is very difficult to estimate or determine an optimal number in ICA for brain functional network extraction, and other work has suggested that multiple model orders contain complementary information [3]. Regarding the analysis of multi-subject fMRI data, the commonly applied group ICA [4] also requires the input of the component number to yield group-level networks for further individual-subject network computation. Indeed, some early work suggests estimating the number of components in ICA using information theoretic criteria (ITC) [5] such as minimum description-length criterion (MDL), Akaike's information criterion (AIC), and Bayesian information criterion (BIC). However, none of these methods works perfectly for the estimation due to complex noise structures [6]. In addition, since there is an assumption of independently and identically distributed samples in ITC, downsampling is often conducted before the estimation, which may cause degradation in the accuracy of the order estimation. In another study, based on the finite memory length and the autoregressive model, entropy-rate-based components number selection methods using ITC were proposed to utilize all available subjects [7]. Due to that different methods could result in diverse estimations, in practice researchers often set an empirical or arbitrary number like 20-40 for low model order and 100-200 for high model order in fMRI studies [1, 8].

Some other studies tried to search the optimal components or the component number by setting varying settings. With the component number ranging from 2 to 100, a study by Kairov ranks the obtained independent components based on their stability and reproducibility in multiple ICA runs (with random initialization) in order to yield a maximally stable model order, however, the method has not been applied to fMRI studies [9]. Kuang et al. estimated functional networks with different model orders and selected the 'best' result that fits well with the reference networks [10].

Recently, a method called Snowball ICA was proposed, which generates seed components by performing ICA on randomly selected subjects' fMRI data and then updates the seed components iteratively by adding different blocks of fMRI data [11], until all subjects' data are used. Although the method does not require the component number as an input, the method may be sensitive to the generation of the seed components and the organization manner of fMRI blocks. Moreover, Snowball ICA is very time-consuming.

So far, how to estimate reliable functional networks using ICA is still a challenging issue in the neuroscience field. In this paper, we propose a model-order free ICA method, named SMART (splitting-merging assisted reliable) ICA, which achieves automatic estimation of reliable independent components (ICs) by combining ICA with splitting and merging clustering. Using simulations, we validate that the components can be perfectly extracted by our method. Using real fMRI data, our method also successfully extracted meaningful subject-specific functional networks.

*Research supported by the National Natural Science Foundation of China (Grant No. 62076157 and 61703253, to YHD).

Yuhui Du is with the School of Computer and Information Technology, Shanxi University, Taiyuan, China (e-mail: duyuhui@sxu.edu.cn).

Xingyu He is with the School of Computer and Information Technology, Shanxi University, Taiyuan, China (e-mail: 1210109650@qq.com).

Vince D Calhoun is with the Tri-Institutional Center for Translational Research in Neuroimaging and Data Science (TReNDS), Georgia State University, Georgia Institute of Technology, Emory University, Atlanta, GA, USA (e-mail: vcalhoun@gsu.edu).

II. OUR METHOD

A. Our SMART (splitting-merging assisted reliable) ICA method

Our new method combines ICA with a clustering technique to yield reliable brain functional networks based on the independent components obtained from ICA runs with different model orders. Our method can be applied to both ICA on individual subject and group ICA on multi-subjects. In the following, we describe how SMART ICA estimates reliable group-level components. Based on this, we extend SMART ICA to enable it to estimate individual-level networks by utilizing our previously proposed group information guided ICA (GIG-ICA) [12, 13] with the guidance of reliable group-level components. Fig. 1 shows the pipeline of our method that primarily includes three steps.

In Step 1, ICA is performed with different numbers of components as the input to obtain the initial group-level ICs that have different network scales. We use $X = (X_1, X_2, \dots, X_n, \dots, X_N)$ to denote the fMRI data of N subjects. $X_n \in R^{T \times V}$, where T $(1, 2, \dots, t)$ represents the number of time points, and V $(1, 2, \dots, v)$ represents the number of voxels within a brain mask. Before ICA, we perform subject-level PCA on each subject's data X_n and then perform a group-level PCA on the concatenation of the reduced data, resulting in a whole matrix D. Then, the Infomax algorithm [14, 15] is implemented to decompose D into k independent components.

$$D \approx A \times S, \tag{1}$$

Here, $A \in R^{k \times k}$ represents the mixing matrix and $S \in R^{k \times V}$ represents the group-level ICs.

We set k to different numbers, thus resulting in a total of g initial group-level ICs, represented by $F \in R^{g \times V}$, that will be utilized to automatically generate reliable group-level ICs.

In Step 2, we cluster those initial group-level ICs by using a community detection method followed by a merging and splitting technique to generate reliable group-level ICs. First, for all initial group-level ICs (i.e. F), a community detection method [16] is performed to cluster the g initial group-level ICs. And then, inspired by previous studies [17], we propose the following rules of splitting and merging to refine the clustering result from the community detection so as to get reliable group-level ICs.

Before introducing the rules, we define that the distance between any two components is calculated by equation (2).

$$d(x, y) = 1 - |corr(x, y)|, \tag{2}$$

where x and y represent any two components, $corr(\cdot)$ represents the Pearson correlation coefficient, and $|\cdot|$ represents absolute value operation. We also determine that a cluster center is defined as the component that is closest to the mean component in the cluster, which can be formulated as (3).

$$\min_{s_i} \{d(s_i, s_{mean})\}, \tag{3}$$

where s_i represents the cluster center and s_{mean} represents the mean of all components in the current cluster.

The rules of splitting:

Process s1: Calculate the average inter-cluster distance d_{mean}.

$$d_{mean} = \frac{2}{z \times (z-1)} \sum_{i=1}^{z} \sum_{j=i+1}^{z} d(s_i, s_j), \tag{4}$$

where s_i and s_j represent the cluster centers, z represents the current cluster number.

Process s2: Calculate the sum (d_{intra}) of the distance between the cluster center and the nearest component (except for itself) as well as the distance between the cluster center and the farthest component for each cluster.

$$d_{intra} = \max_{x_p} \{d(s_i, x_p)\} + \min_{x_p} \{d(s_i, x_p)\}, s.t. x_p \neq s_i \tag{5}$$

where s_i represents the cluster center, and x_p represents any component in this cluster.

Process s3: A cluster should be split if $d_{intra} > d_{mean}/2$. Then, another community detection is conducted to split this cluster to update the clustering result.

Process s4: Update the cluster center and the number of clusters.

Process s5: If the number of the cluster remains the same, stop the splitting process, otherwise, return to Process s1.

The rules of merging:

Process m1: Calculate the average inter-cluster distance d_{mean} using formula (4).

Process m2: Calculate the distance (d_{inter}) between any two cluster centers for all clusters using formula (2).

Process m3: Two clusters should be merged if $d_{inter} < d_{mean}/2$.

Process m4: Update the cluster centers and the number of clusters.

Process m5: If the number of the cluster remains the same, stop the merging process, otherwise, return to Process m1.

The operation of splitting and merging is iteratively conducted until the cluster labels of the initial group-level ICs don't change. As such, the initial group-level ICs are grouped into different clusters, whose cluster centers represent the reliable group-level ICs.

In Step 3, we take reliable group-level ICs as reference information to perform GIG-ICA [13, 18] to obtain individual-level ICs and related time series. After that, automatic denoising is done to extract meaningful functional networks.

B. Validation using simulation

The simulation data to evaluate our method are generated via the SimTB toolbox [19]. Two groups (group A and group B, each group contains 15 subjects) are generated. Each subject's data are obtained using eight spatial components (SMs) and their related time series (with 150 time points). Among the eight SMs, six are common across the two groups (A and B) and two are unique for each group. Note: random spatial transition, rotation, and deformation are added for SMs, and noises are simulated as well.

For the simulation data, k is set from 5 to 15 in Step 1, so g (110) initial group-level ICs are computed. In Step 3, we propose to denoise ICs by measuring the similarity between

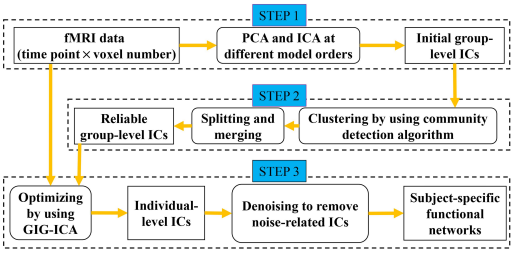

Fig. 1. The pipeline of SMART ICA.

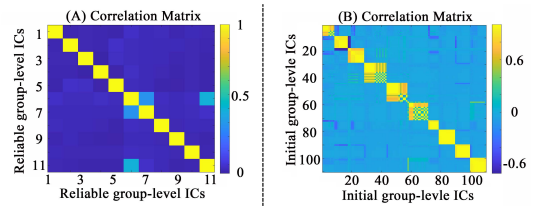

Fig. 2. Results of reliable group-level ICs. (A) Correlation matrix of 12 reliable group-level ICs. (B) Correlation matrix of 110 initial group-level ICs after the sorting according to the cluster labels.

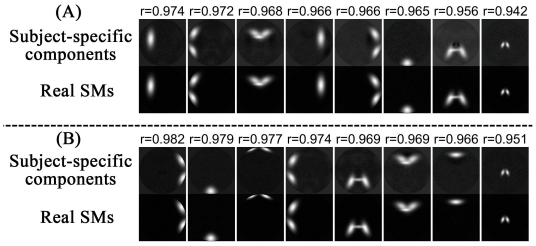

Fig. 3. The subject-specific components for one example subject from group A (see Fig. 3(A)) and group B (see Fig. 3(B)), respectively. In both (A) and (B), we include correlation r between the extracted subject-specific ICs and real SMs (the first line), the subject-specific ICs (the second line), and the real SMs (the third line).

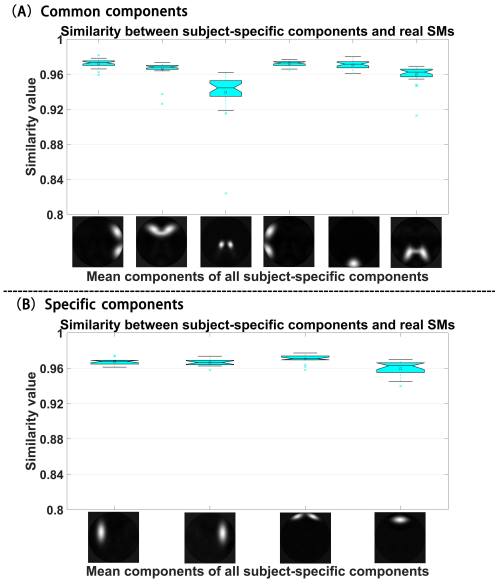

Fig. 4. The similarity between subject-specific ICs and real SMs of all components across all subjects. (A) displays the similarity of six common ICs (note: each IC is included in the data of all 30 subjects from group A and group B). (B) displays the similarity of four subject-unique ICs. The first two come from group A and the last two come from group B.

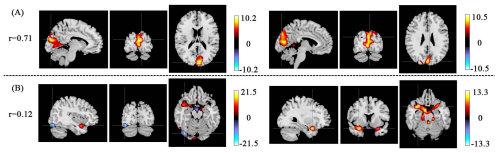

Fig. 5. Some subject-specific functional networks estimated by SMART ICA using real fMRI data. In (A), we show two networks with the greatest mean similarity ($r = 0.71$) across subjects. In (B), two networks with the lowest similarity ($r = 0.12$) are shown. Note: the corresponding left and right networks were estimated using the same reliable group-level IC.

level IC reflected by the number of Maximally Stable Extremal Regions (MSER) [20]. Here, if the similarity between any two ICs is greater than 0.5, we average the two ICs as a new IC. For each IC, if the MSER number is less than 150, the IC is considered as one network, otherwise, it is taken as noise. It's worth noting that our method is insensitive to the two parameters.

To show the results of simulation data, we match the extracted subject-specific components and real SMs of all subjects according to the correlation r between them and display the real SMs and extracted subject-specific components of two subjects that come from group A and group B, respectively. In addition, the similarity between subject-specific functional networks and real SMs of all components across all subjects is displayed by using boxplots.

C. Validation using real fMRI

To verify the effectiveness of SMART ICA, we analyze fMRI data from 25 healthy controls. Here, k is set to include both low (20, 25, and 30) and high (75 and 100) model orders in Step 1. In Step 3, denoising can be conducted using a similar way to that in simulation-based experiments or using a toolbox such as Noisecloud in GIFT [21]. For the subject-specific components estimated using the same reliable group-level IC, we calculate their similarity between different subjects to observe the consistency and unique properties of functional networks. Meanwhile, we show some example subject-specific networks.

III. Result

A. Results using simulation

By clustering 110 initial group-level ICs, 11 reliable group-level ICs were obtained. Fig. 2(A) shows the correlation coefficients of 11 reliable group-level ICs, meaning that each reliable group-level IC has a unique pattern. Fig. 2(B) shows the correlation coefficient matrix of 110 initial group-level ICs, sorted according to clustering labels. It is seen that there are high intra-cluster similarity and low inter-cluster similarity in Fig. 2(B), indicating that our method performs effective clustering on the initial group-level ICs.

After denoising processing on the individual-subject ICs, eight ICs were preserved for each subject. For each group

(group A and B), we show the obtained ICs for one example subject in Fig. 3. We found that both the common and unique components were perfectly extracted, and the similarity between the estimated components and the real SMs is very high for the two subjects. The similarity between subject-specific functional networks and real SMs of all components across all subjects is summarized in Fig. 4 using boxplots. It is seen that the overall accuracy of the estimated components is high.

B. Results using fMRI data

By applying our SMART ICA on real fMRI data, 103 reliable group-level ICs were obtained based on 250 initial group-level ICs. Among the 103 ICs, 3 ICs were retained from the results of k = 20, 5 ICs were retained from the results of k = 25, 11 ICs were retained from the results of k = 30, 23 ICs were retained from the results of k = 75, and 61 ICs were retained from the results of k = 100, supporting that our method can simultaneously take advantage of networks with different scales. When the reliable group-level ICs came from low model order, the mean percentage of associated ICs that belonged to the same cluster and came from the lower order models was 71.8%. When the reliable group-level ICs came from high model order, the percentage of associated ICs that belonged to the same cluster and came from the high order models was 90.5%. Fig. 5 displays four subject-specific networks that were extracted by SMART ICA using real data. Here, we show two subject-specific networks that were most similar and two subject-specific networks that were most different. The results show that our method not only can find biologically corresponding functional networks but also can explore unique networks among different subjects.

IV. Discussion and Conclusion

ICA has been widely applied to the analysis of brain functional networks using fMRI. However, how to determine the model order is a difficult problem, although there have been some efforts [11]. In this paper, we propose a method, called SMART ICA, that can automatically identify subject-specific components, without the need of setting a specific number of components. In our method, an advanced clustering technique is proposed to integrate the ICs from varying model-order settings so that the functional networks with multi-scales can be preserved.

Our method performs well for both simulations and real fMRI data. Regarding the two groups of simulated data, the number of subject-specific components that are obtained by our method is the same as the number of real SMs. Our result shows that SMART ICA can accurately estimate the components even when subjects have unique components. Using fMRI data, we not only found that the visual networks show the greatest similarity between different subjects but also revealed that some unique networks relate to higher-level functions (e.g., parahippocampal and temporal pole). Taken together, our method is promising for rapidly promoting the application of ICA on fMRI analysis, as our method can simultaneously capture components estimated well at different scales (both low and high model orders).

Acknowledgment

This work was supported by the National Natural Science Foundation of China (Grant No. 62076157 and 61703253, to YHD) and the National Institutes of Health grant R01MH123610 to VDC. We acknowledge the contribution of all participants in this project.

Reference

[1] H. Xiong *et al.*, "Altered Default Mode Network and Salience Network Functional Connectivity in Patients with Generalized Anxiety Disorders: An ICA-Based Resting-State fMRI Study," *Evidence-Based Complementary and Alternative Medicine,* vol. 2020, 2020.

[2] M. Ohta *et al.*, "Structural equation modeling approach between salience network dysfunction, depressed mood, and subjective quality of life in schizophrenia: an ICA resting-state fMRI study," *Neuropsychiatric disease and treatment,* vol. 14, pp. 1585, 2018.

[3] A. Iraji *et al.*, "Multi-spatial scale dynamic interactions between functional sources reveal sex-specific changes in schizophrenia," *Network Neuroscience,* pp. 1-48, 2021.

[4] V. D. Calhoun and N. de Lacy, "Ten Key Observations on the Analysis of Resting-state Functional MR Imaging Data Using Independent Component Analysis," *Neuroimaging clinics of North America,* vol. 27, no. 4, pp. 561-579, 2017.

[5] Y. O. Li *et al.*, "Estimating the number of independent components for functional magnetic resonance imaging data," (in eng), *Human brain mapping,* vol. 28, no. 11, pp. 1251-66, 2007.

[6] M. Hui *et al.*, "An empirical comparison of information-theoretic criteria in estimating the number of independent components of fMRI data," *PloS one,* vol. 6, no. 12, pp. e29274, 2011.

[7] G.-S. Fu *et al.*, "Likelihood estimators for dependent samples and their application to order detection," *IEEE transactions on signal processing,* vol. 62, no. 16, pp. 4237-4244, 2014.

[8] K. S. Gopinath *et al.*, "Exploring brain mechanisms underlying Gulf War Illness with group ICA based analysis of fMRI resting state networks," *Neuroscience letters,* vol. 701, pp. 136-141, 2019.

[9] U. Kairov *et al.*, "Determining the optimal number of independent components for reproducible transcriptomic data analysis," *BMC genomics,* vol. 18, no. 1, pp. 1-13, 2017.

[10] L.-D. Kuang *et al.*, "Model order effects on ICA of resting-state complex-valued fMRI data: application to schizophrenia," *Journal of neuroscience methods,* vol. 304, pp. 24-38, 2018.

[11] G. Hu *et al.*, "Snowball ICA: A Model Order Free Independent Component Analysis Strategy for Functional Magnetic Resonance Imaging Data," *Frontiers in neuroscience,* vol. 14, pp. 1005, 2020.

[12] Y. Du *et al.*, "NeuroMark: An automated and adaptive ICA based pipeline to identify reproducible fMRI markers of brain disorders," *NeuroImage: Clinical,* vol. 28, pp. 102375, 2020.

[13] Y. Du and Y. Fan, "Group information guided ICA for fMRI data analysis," *Neuroimage,* vol. 69, pp. 157-197, 2013.

[14] S.-i. Amari *et al.*, "A new learning algorithm for blind signal separation," in *Advances in neural information processing systems*, Morgan Kaufmann Publishers, pp. 757-763, 1996.

[15] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural computation,* vol. 7, no. 6, pp. 1129-1159, 1995.

[16] E. A. Leicht and M. E. Newman, "Community structure in directed networks," *Physical review letters,* vol. 100, no. 11, pp. 118703, 2008.

[17] J. Lei *et al.*, "Robust K-means algorithm with automatically splitting and merging clusters and its applications for surveillance data," *Multimedia Tools and Applications,* vol. 75, no. 19, pp. 12043-12059, 2016.

[18] Y. Du *et al.*, "Identifying dynamic functional connectivity biomarkers using GIG‐ICA: Application to schizophrenia, schizoaffective disorder, and psychotic bipolar disorder," *Human brain mapping,* vol. 38, no. 5, pp. 2683-2708, 2017.

[19] E. B. Erhardt *et al.*, "SimTB, a simulation toolbox for fMRI data under a model of spatiotemporal separability," *Neuroimage,* vol. 59, no. 4, pp. 4160-4167, 2012.

[20] J. Matas *et al.*, "Robust wide-baseline stereo from maximally stable extremal regions," *Image and vision computing,* vol. 22, no. 10, pp. 761-767, 2004.

[21] V. Sochat *et al.*, "A robust classifier to distinguish noise from fMRI independent components," *PLoS One,* vol. 9, no. 4, pp. e95493, 2014.